Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 77 (2007) 1258–1268
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Extensions of the Markov chain marginal bootstrap ☆

Masha Kocherginsky[a], Xuming He[b,*]

[a]Department of Health Studies, University of Chicago, USA
[b]Department of Statistics, University of Illinois, Urbana-Champaign, USA

Available online 16 March 2007

Abstract

The Markov chain marginal bootstrap (MCMB) was introduced by He and Hu [2002. Markov chain marginal bootstrap. J. Amer. Statist. Assoc. 97(459) (2002) 783–795] as a bootstrap-based method for constructing confidence intervals or regions for a wide class of *M*-estimators in linear regression and maximum likelihood estimators in certain parametric models. In this article we discuss more general applications of MCMB-**A**, an extension of the MCMB algorithm, which was first proposed in Kocherginsky et al. [2005. Practical confidence intervals for regression quantiles. J. Comput. Graphical Statist. 14, 41–55] for quantile regression models. We also present a further extension of the MCMB algorithm, the **B**-transformation, which is a transformation of the estimating equations, aiming to broaden the applicability of the MCMB algorithm to general estimating equations that are not necessarily likelihood-based. We show that applying the **A**- and **B**-transformations jointly enables the MCMB algorithm to be used for inference related to a very general class of estimating equations. We illustrate the use of the MCMB-**AB** algorithm with a nonlinear regression model with heteroscedastic error distribution.

Keywords: Bootstrap; General estimating equations; Confidence intervals; Transformation; Nonlinear regression

1. Introduction

The key idea of the bootstrap is to resample from the empirical distribution of the observed data and subsequently to estimate the quantities of interest (e.g., standard errors and confidence regions) from the bootstrap samples. The bootstrap methods are particularly useful when the asymptotic approximation is hard to compute as it foregoes "the long-winded and error-prone analytical calculation" (Davison and Hinkley, 1997). The bootstrap can, however, be computationally expensive and even prohibitive for large-dimensional problems.

Regression models are among the most important and frequent applications of the bootstrap. Consider a regression model

$$y_i = f(\mathbf{x}_i, \boldsymbol{\theta}) + e_i, \tag{1}$$

☆The research is partially supported by the National Science Foundation Award DMS-0604229 and a National Security Agency Grant.
*Corresponding author.
E-mail addresses: mkocherg@uchicago.edu (M. Kocherginsky), x-he@uiuc.edu (X. He).

0167-7152/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2007.03.010

where the response variable Y is related to the independent variable $\mathbf{X}$ (which can be a matrix) via a function f and a vector of unknown parameters $\boldsymbol{\theta}$. The residual and the paired bootstrap are two common methods for bootstrap estimation of $\boldsymbol{\theta}$ (Efron and Tibshirani, 1998). The residual bootstrap works well when the errors are exchangeable in distribution, while the paired bootstrap is based on the assumption that the observations come from a multivariate distribution $F(\mathbf{X}, \mathbf{Y})$. Either residuals or pairs of observations $(\mathbf{x}_i, y_i)$ are resampled K times, respectively, and the bootstrap samples $(\mathbf{x}_i^*, y_i^*)$ are used to produce estimates $\hat{\boldsymbol{\theta}}^{*(1)}, \ldots, \hat{\boldsymbol{\theta}}^{*(K)}$ of $\boldsymbol{\theta}$. The paired bootstrap makes no assumption about error variance homogeneity, and is therefore more robust.

An alternative method, which resamples the estimating equations instead of resampling the data directly, was proposed by Hu and Zidek (1995). For a linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$, where $\boldsymbol{\theta} \in \mathbb{R}^p$, $\mathrm{Cov}(\mathbf{e}) = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, the normal estimating equations for the ordinary least-squares estimate $\hat{\boldsymbol{\theta}}$ solve $\sum_{i=1}^n \mathbf{x}_i r_i = 0$, where r_i's are the residuals. This can be reexpressed as

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \sum_{i=1}^{n} \mathbf{x}_i e_i. \tag{2}$$

The bootstrap version of (2) becomes

$$\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \sum_{i=1}^{n} \mathbf{z}_i^*, \tag{3}$$

where $\mathbf{z}_i^*$ is a bootstrap sample from $(\mathbf{x}_1 r_1, \ldots, \mathbf{x}_n r_n)$. This method is robust against nonhomogeneity of the second and higher moments.

A common feature of all bootstrap methods is that the same estimation procedure is applied to each of the bootstrapped data sets, making the bootstrap computationally expensive, especially for moderate to large p. For example, the number of operations for the paired bootstrap using the standard least-squares algorithm is bounded below by Knp^2, where p is the number of parameters and K is the number of bootstrap samples.

He and Hu (2002) introduced a new bootstrap method called the Markov chain marginal bootstrap (MCMB), which reduces the computational complexity of the bootstrap by resampling the marginal estimating equations p times at each bootstrap step k to sequentially solve p one-dimensional equations instead of the p-dimensional system of equations. It uses a partially updated vector of parameter estimates

$$\hat{\boldsymbol{\theta}}^{(k,k-1)} = \left(\hat{\theta}_1^{(k)}, \ldots, \hat{\theta}_{j-1}^{(k)}, \hat{\theta}_j^{(k)}, \hat{\theta}_{j+1}^{(k-1)}, \ldots, \hat{\theta}_p^{(k-1)}\right)$$

to obtain the most current estimate $\hat{\theta}_j^{(k)}$. In Kocherginsky et al. (2005), we applied the MCMB algorithm to quantile regression models (Koenker and Bassett, 1978) and showed that it is a fast and reliable resampling method for problems with large n and p. We also proposed the MCMB-**A** algorithm, where an affine transformation **A** of the parameter space is used to alleviate the problem of possible high autocorrelation of the MCMB sequences. In the MCMB-**A** algorithm the **A**-transformation involves implementing the original MCMB algorithm on a different parameter scale using a transformed parameter vector. While in that article we discussed the application of the **A**-transformation in quantile regression, the autocorrelation of MCMB sequences is an even more common and persistent problem in nonlinear regression models.

In the next three sections we review the MCMB algorithm, discuss the MCMB-**A** method, and describe a further extension of the MCMB algorithm, the **B**-transformation, which is a transformation of the estimating equations, aiming to broaden the applicability of the MCMB algorithm to general estimating equations. With both **A** and **B**-transformations, we show in Section 5 that the MCMB algorithm may be used for inference related to a very general class of estimating equations. We discuss the relationship between the MCMB algorithm and Markov chain Monte Carlo (MCMC) algorithms in Section 6. We then give illustrative examples in Section 7 to show some good properties of the MCMB method, followed by some further remarks in Section 8.

2. Review of MCMB

Let $Y_1, \ldots, Y_n$ be a sequence of independent random vectors, and $\boldsymbol{\theta} \in \mathbb{R}^p$ be an unknown p-dimensional parameter. Suppose that $\boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta})$ are $\mathbb{R}^p$-valued functions of the unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^p$ and the observations Y_i. Assume that the expected value of $\boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta})$'s is the p-dimensional vector of 0's, i.e., $E_\theta[\boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta})] = \mathbf{0}$. Consider $\hat{\boldsymbol{\theta}}$, a solution of the unbiased estimation equation

$$S(\mathbf{Y}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta}) = \mathbf{0}, \tag{4}$$

where $\mathbf{Y}$ denotes the data, and the subscript n on the estimator $\hat{\boldsymbol{\theta}}$ is suppressed. Using this notation, $S(\mathbf{Y}, \boldsymbol{\theta})$ is a p-dimensional vector. Its jth element is $S_j(\mathbf{Y}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \psi_{ij}(Y_i, \boldsymbol{\theta})$, $j = 1, \ldots, p$, and $\boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta})$ is a vector $(\psi_{i1}(Y_i, \boldsymbol{\theta}), \ldots, \psi_{ip}(Y_i, \boldsymbol{\theta}))^{\mathrm{T}}$. Assume that a decomposition $\boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta}) = \mathbf{a}_i z_i$ exists, so that $\mathbf{a}_i$'s are vectors or matrices independent of $\mathbf{Y}$, and z_i's are random variables depending on $\mathbf{Y}$ and $\boldsymbol{\theta}$. For example, in a linear regression model $Y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta} + e_i$, the MCMB can be carried out with $\mathbf{a}_i = \mathbf{x}_i$ and $z_i = e_i$. The $\mathbf{a}_i$'s and z_i's are used as a matter of convenient notation to indicate the quantities that are to be resampled; they do not provide a unique decomposition, allowing for different resampling schemes. Note that $\hat{\boldsymbol{\theta}}$ is an M-estimator. Typically, M-estimators (Huber, 1964) minimize $\sum_{i=1}^{n} \rho(y_i, \boldsymbol{\theta})$ for an objective function ρ, often chosen to be a symmetric and continuous differentiable function. If ρ is differentiable with $\rho' = \psi$, then the M-estimator is defined as the solution of $\sum_{i=1}^{n} \boldsymbol{\psi}(y_i, \boldsymbol{\theta}) = \mathbf{0}$.

The conventional bootstrap solves (4) for each bootstrap sample, which is obtained with residual or paired bootstrap, or by sampling directly from the estimating equations. Solving for $\hat{\boldsymbol{\theta}}$ involves solving the p-dimensional system of Eq. (4) a large number of times. The MCMB instead solves the sequences of one-dimensional equations using the following algorithm:

1. Initialize $\hat{\boldsymbol{\theta}}^{(0)} = \hat{\boldsymbol{\theta}}$ and $k = 1$.
2. (a) Set $j = 1$.
   (b) Draw a bootstrap sample $\{z_{1j}^{*(k)}, \ldots, z_{nj}^{*(k)}\}$ from $\{z_1, \ldots, z_n\}$ for each $j = 1, \ldots, p$.
3. (a) Solve for $\hat{\theta}_j^{(k)}$

$$S_j\left(\mathbf{Y}; \hat{\theta}_1^{(k)}, \ldots, \hat{\theta}_{j-1}^{(k)}, \boxed{\hat{\theta}_j^{(k)}}, \hat{\theta}_{j+1}^{(k-1)}, \ldots, \hat{\theta}_p^{(k-1)}\right) = S_j^{*(k)}\left(\mathbf{Y}; \hat{\boldsymbol{\theta}}^{(0)}\right), \tag{5}$$

   where $(\hat{\theta}_1^{(k)}, \ldots, \hat{\theta}_{j-1}^{(k)}, \hat{\theta}_{j+1}^{(k-1)}, \ldots, \hat{\theta}_p^{(k-1)})$ is the most current estimate of $\hat{\boldsymbol{\theta}}$ without the jth component, and $S_j^{*(k)}(\mathbf{Y}; \hat{\boldsymbol{\theta}}^{(0)}) = \sum_{i=1}^{n} \mathbf{a}_i \mathbf{z}_{ij}^{*(k)}$.
   (b) Return to 2(a) with $j = j + 1$, if $j < p$.
4. Increase k by 1 and go to step 2a or stop if k has reached a pre-specified level K.

This algorithm results in a sequence of solutions $\hat{\boldsymbol{\theta}}^{(1)}, \ldots, \hat{\boldsymbol{\theta}}^{(K)}$, which by construction is a Markov chain. At the kth iteration the MCMB algorithm solves p one-dimensional equations for $\hat{\theta}_j^{(k)}$, $j = 1, \ldots, p$ using the already updated components $\hat{\theta}_1^{(k)}, \ldots, \hat{\theta}_{j-1}^{(k)}$ together with the values remaining from the previous iteration $\hat{\theta}_{j+1}^{(k-1)}, \ldots, \hat{\theta}_p^{(k-1)}$.

3. Discussion on MCMB-A

Recall that MCMB solves the unbiased estimation equations (4) for $\boldsymbol{\theta}$. If the covariance matrix of $S(\mathbf{Y}, \boldsymbol{\theta})$ is nearly singular, the resulting MCMB sequences will be strongly autocorrelated. We propose a transformation of the parameter space, which we call the **A**-transformation, which "decorrelates" the MCMB sequences. Accordingly, consider a $p \times p$ matrix $\mathbf{A}$ and a vector $\tilde{\boldsymbol{\theta}}$ such that

$$\boldsymbol{\theta} = \mathbf{A}\tilde{\boldsymbol{\theta}},$$

and consequently $\tilde{\boldsymbol{\theta}} = \mathbf{A}^{-1}\boldsymbol{\theta}$. The transformed estimating equations are

$$\tilde{S}(\mathbf{Y}, \tilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \frac{\partial}{\partial \tilde{\boldsymbol{\theta}}} \rho(\mathbf{y}_i, \mathbf{A}\tilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\theta}}{\partial \tilde{\boldsymbol{\theta}}} \frac{\partial}{\partial \boldsymbol{\theta}} \rho(y_i, \mathbf{A}\tilde{\boldsymbol{\theta}}) = \mathbf{A} \cdot S(\mathbf{Y}, \boldsymbol{\theta}).$$

In order to "decorrelate" the sequences of MCMB solutions, the covariance matrix of $\tilde{S}(\mathbf{Y}, \tilde{\boldsymbol{\theta}})$ should be aimed at the identity matrix:

$$\mathrm{Cov}(\tilde{S}(\mathbf{Y}, \tilde{\boldsymbol{\theta}})) = \mathbf{A}\,\mathrm{Cov}(S(\mathbf{Y}, \boldsymbol{\theta}))\,\mathbf{A}^{\mathrm{T}} = \mathbf{I}.$$

Solving this equation results in

$$\mathbf{A} = \mathrm{Cov}(S(\mathbf{Y}, \boldsymbol{\theta}))^{-1/2}. \tag{6}$$

Note that $\mathbf{A}$ exists since $\mathrm{Cov}(S(\mathbf{Y}, \boldsymbol{\theta}))$ is symmetric and is assumed to be invertible, and its square root can be found using the singular value decomposition. The covariance matrix $\mathrm{Cov}(S(\mathbf{Y}, \boldsymbol{\theta}))$ can be estimated by

$$\mathrm{Cov}(S(\mathbf{Y}, \boldsymbol{\theta})) \approx \frac{1}{n}\{\boldsymbol{\psi}(\mathbf{Y}, \hat{\boldsymbol{\theta}})^{\mathrm{T}}\, \boldsymbol{\psi}(\mathbf{Y}, \hat{\boldsymbol{\theta}})\},$$

where $\hat{\boldsymbol{\theta}}$ is the initial estimate of $\boldsymbol{\theta}$, and $\boldsymbol{\psi}(\mathbf{Y}, \hat{\boldsymbol{\theta}})$ is an $n \times p$ matrix with elements $\psi_{ij}(y_i, \hat{\boldsymbol{\theta}})$.

Let $S(\mathbf{Y}, \boldsymbol{\theta})$ be such that $\partial/\partial\boldsymbol{\theta}\ S(\mathbf{Y}, \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, and the matrix $E(\partial/\partial\boldsymbol{\theta}\ S(\mathbf{Y}, \boldsymbol{\theta}))$ is nonsingular. If $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$ and solves $S(\mathbf{Y}, \boldsymbol{\theta}) = \mathbf{0}$, then, from the Taylor expansion,

$$\text{A-Var}(\hat{\boldsymbol{\theta}}) = \left(E\frac{\partial}{\partial \boldsymbol{\theta}} S(\mathbf{Y}, \boldsymbol{\theta})\right)^{-1} \mathrm{Var}(S(\mathbf{Y}, \boldsymbol{\theta})) \left(E\frac{\partial}{\partial \boldsymbol{\theta}} S(\mathbf{Y}, \boldsymbol{\theta})\right)^{-1},$$

where $\text{A-Var}(\hat{\boldsymbol{\theta}})$ is the asymptotic variance of $\hat{\boldsymbol{\theta}}$. Furthermore, if $\mathrm{Var}(S(\mathbf{Y}, \boldsymbol{\theta})) = \pm E(\partial/\partial\boldsymbol{\theta}\ S(\mathbf{Y}, \boldsymbol{\theta}))$, then $\text{A-Var}(\hat{\boldsymbol{\theta}}) = \mathrm{Var}(S(\mathbf{Y}, \boldsymbol{\theta}))^{-1}$. Also, it follows from (6) that

$$\begin{aligned}\text{A-Var}(\hat{\tilde{\boldsymbol{\theta}}}) &= \mathbf{A}^{-1}\left(\frac{\partial}{\partial \boldsymbol{\theta}} S(\mathbf{Y}, \boldsymbol{\theta})\right)^{-1} \mathrm{Var}(S(\mathbf{Y}, \boldsymbol{\theta})) \left(\frac{\partial}{\partial \boldsymbol{\theta}} S(\mathbf{Y}, \boldsymbol{\theta})\right)^{-1} \mathbf{A}^{-1} \\ &= \mathrm{Var}(S(\mathbf{Y}, \boldsymbol{\theta}))^{1/2} \mathrm{Var}(S(\mathbf{Y}, \boldsymbol{\theta}))^{-1} \mathrm{Var}(S(\mathbf{Y}, \boldsymbol{\theta}))^{1/2} = \mathbf{I}\end{aligned}$$

when $E(\partial/\partial\boldsymbol{\theta}\ S(\mathbf{Y}, \boldsymbol{\theta})) = \mathrm{Var}(S(\mathbf{Y}, \boldsymbol{\theta}))^{-1}$. Thus, the $\tilde{\theta}_j$'s are uncorrelated, and the variance of $\hat{\tilde{\theta}}_j$ is 1. The latter is a useful property, since it provides the scale for the variance of $\tilde{\boldsymbol{\theta}}$ when the initial estimate may not be accurate.

In practice, any consistent estimate of $\mathbf{A}$ up to a scale factor may be used for the same purpose of decorrelating the MCMB sequence. For this reason, the $\mathbf{A}$-transformation for a linear model can be used in the following simple way:

$$\mathbf{X}\boldsymbol{\theta} = \mathbf{X}\mathbf{A}\tilde{\boldsymbol{\theta}} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\theta}},$$

where $\tilde{\mathbf{X}} = \mathbf{X}\,\mathbf{A}$, and $r_i = y_i - \mathbf{X}\boldsymbol{\theta} = y_i - \tilde{\mathbf{X}}\tilde{\boldsymbol{\theta}}$. In this special case, the MCMB algorithm can be applied directly on $\tilde{\boldsymbol{\theta}}$ scale if the design matrix $\mathbf{X}$ is replaced by $\tilde{\mathbf{X}}$ in the beginning, and transformed back in the end. The transformation between $\tilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$ becomes unnecessary at every step.

4. Introduction to MCMB-B

He and Hu (2002) provided the conditions under which the distribution of $\hat{\boldsymbol{\theta}}^{(k)}$ resulting from the MCMB algorithm is an asymptotically valid approximation to that of $\hat{\boldsymbol{\theta}}$. Again, we consider an estimator defined by (4). The derivations in He and Hu (2002) show that if the following conditions hold, the MCMB sequence does approximate the distribution of $\hat{\boldsymbol{\theta}}$ as desired.

(C1) $\partial\boldsymbol{\psi}_i(y_i, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is continuously differentiable in $\boldsymbol{\theta}$;
(C2) $n^{-1}\sum_{i=1}^{n}(\partial^2/\partial\boldsymbol{\theta}^2)\boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta})$ is uniformly bounded in n, and $\sum_{i=1}^{n} E|\boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta})|^3 = \mathcal{O}(n)$ as $n \to \infty$;
(C3) both $n^{-1}\sum_{i=1}^{n}[\boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta})\ \boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta})^{\mathrm{T}}]$ and $-n^{-1}\sum_{i=1}^{n}(\partial/\partial\boldsymbol{\theta})\boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta})$ converge to the information matrix $D(\boldsymbol{\theta}) > 0$ as $n \to \infty$.

Condition (C3) is essential for $\sqrt{n}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{(k)})$ to be an approximate multiple AR(1) process. This condition holds if (4) is a likelihood equation and Y_i are i.i.d. with $\boldsymbol{\psi}_i = \boldsymbol{\psi}$ for all i, where D is the Fisher information matrix. In general, however we cannot expect (C3) to hold.

We propose a transformation of the estimating equations which ensures that (C3) holds. We call it the **B**-transformation, as it "broadens" the applicability of the MCMB. For any estimator $\hat{\boldsymbol{\theta}}$ which is the solution of $\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta}) = S(Y, \boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{\psi}_i(Y_i, \boldsymbol{\theta}) = \mathbf{0}$, where $\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})$ is a p-dimensional vector consider a linear transformation of the estimating equations

$$\tilde{\boldsymbol{\psi}}(\mathbf{Y}, \boldsymbol{\theta}) = \mathbf{B}\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta}),$$

where $\mathbf{B}$ is a $p \times p$ nonsingular matrix and $\tilde{\boldsymbol{\psi}}(\mathbf{Y}, \boldsymbol{\theta})$ is a $p \times 1$ vector. Multiplying the original set of equations by a nonsingular matrix transforms the marginal equations, but does not affect the solution $\hat{\boldsymbol{\theta}}$. The new equations have mean and variance

$$E\frac{\partial}{\partial \boldsymbol{\theta}}\tilde{\boldsymbol{\psi}}(\mathbf{Y}, \boldsymbol{\theta}) = \mathbf{B}\, E\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta}), \quad \mathrm{Var}(\tilde{\boldsymbol{\psi}}(\mathbf{Y}, \boldsymbol{\theta})) = \mathbf{B}\,\mathrm{Var}(\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta}))\,\mathbf{B}^{\mathrm{T}}. \tag{7}$$

In order for the condition (C3) to hold under the transformed equations, we require

$$-E_{\theta}\frac{\partial \tilde{\boldsymbol{\psi}}(\mathbf{Y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathrm{Var}(\tilde{\boldsymbol{\psi}}(\mathbf{Y}, \boldsymbol{\theta})). \tag{8}$$

Substituting (7) into (8) and solving for $\mathbf{B}$ results in

$$\mathbf{B} = -E_{\theta}\left[\frac{\partial \boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^{\mathrm{T}} \cdot [\mathrm{Var}(\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta}))]^{-1},$$

which ensures that (C3) holds for the new equations. Note that if (C3) holds for the original equations $\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})$ then $\mathbf{B} = \mathbf{I}$ and $\tilde{\boldsymbol{\psi}}(\mathbf{Y}, \boldsymbol{\theta}) = \boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})$.

The quantities $E_{\theta}[\partial/\partial\boldsymbol{\theta}\ \boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})]$ and $\mathrm{Var}(\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta}))$ can be obtained using the initial estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Also note that (C3) needs to hold only for the true parameter $\boldsymbol{\theta}$. If $\mathbf{B}$ is evaluated at $\hat{\boldsymbol{\theta}}$, which is a consistent estimate of $\boldsymbol{\theta}$, (8) holds asymptotically. We refer to this estimation procedure as MCMB-**B**.

5. Journey to MCMB-AB

Although the **B**-transformation enforces (C3), it does nothing to eliminate the autocorrelation of the resulting MCMB sequences. We therefore propose to use both **A**- and **B**-transformations jointly to address both issues simultaneously, and call the resulting algorithm MCMB-**AB**.

Let matrices $\mathbf{P}$ and $\mathbf{Q}$ be

$$\mathbf{P} = E\left[\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})\right]^{\mathrm{T}} \quad \text{and} \quad \mathbf{Q} = E[\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})^{\mathrm{T}}]. \tag{9}$$

The possible dependence of $\mathbf{P}$ and $\mathbf{Q}$ on n will be suppressed. Also, let $\boldsymbol{\theta} = \mathbf{A}\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\psi}}(\mathbf{Y}, \tilde{\boldsymbol{\theta}}) = \mathbf{A}\boldsymbol{\psi}(\mathbf{Y}, \mathbf{A}\tilde{\boldsymbol{\theta}}) = \mathbf{0}$, as before. The new estimating equations, which result from applying the **B**-transformation after the **A**-transformation, become

$$\mathbf{BA}\boldsymbol{\psi}(\mathbf{Y}, \mathbf{A}\tilde{\boldsymbol{\theta}}) = \mathbf{0}. \tag{10}$$

We choose $\mathbf{B}$ so that condition (C3) holds for the new equations in (10), that is

$$E\left[\frac{\partial}{\partial \tilde{\boldsymbol{\theta}}}(\mathbf{BA}\boldsymbol{\psi}(\mathbf{Y}, \mathbf{A}\tilde{\boldsymbol{\theta}}))\right] = -E[(\mathbf{BA}\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})) \cdot (\mathbf{BA}\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta}))^{\mathrm{T}}]. \tag{11}$$

The expression on the left side of (11) simplifies to $\mathbf{B}\ \mathbf{A}\ E[(\partial/\partial\boldsymbol{\theta})\boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\theta})] \cdot \partial\boldsymbol{\theta}/\partial\tilde{\boldsymbol{\theta}} = \mathbf{B}\ \mathbf{A}\ \mathbf{P}\ \mathbf{A}$, and on the right side simplifies to $-\mathbf{B}\,\mathbf{A}\,\mathbf{Q}\,\mathbf{A}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}$. Equating the two and solving for $\mathbf{B}$ results in

$$\mathbf{B} = -\mathbf{APQ}^{-1}\mathbf{A}^{-1}. \tag{12}$$

Next, we choose $\mathbf{A}$ to eliminate the autocorrelation, so that

$$\mathrm{Var}(\mathbf{BA}\boldsymbol{\psi}(\mathbf{Y},\boldsymbol{\theta})) = \mathbf{I}.$$

Substituting $\mathbf{B}$ from (12), and solving for $\mathbf{A}$ results in

$$\mathbf{A} = (n \cdot \mathbf{PQ}^{-1}\mathbf{P})^{-1/2}. \tag{13}$$

Note that if $\tilde{\boldsymbol{\theta}}$ is the solution of $\mathbf{B}\tilde{\boldsymbol{\psi}}(\mathbf{Y},\tilde{\boldsymbol{\theta}}) = \mathbf{0}$, then the asymptotic covariance is $\text{A-Var}(\hat{\tilde{\boldsymbol{\theta}}}) = \mathbf{I}$, since

$$\begin{aligned}\text{A-Var}(\hat{\tilde{\boldsymbol{\theta}}}) &= E\left(\frac{\partial}{\partial\tilde{\boldsymbol{\theta}}}\mathbf{BA}S(\mathbf{Y},\boldsymbol{\theta})\right)^{-1} \cdot \mathrm{Var}(\mathbf{BA}S(\mathbf{Y},\boldsymbol{\theta})) \cdot E\left(\frac{\partial}{\partial\tilde{\boldsymbol{\theta}}}\mathbf{BA}S(\mathbf{Y},\boldsymbol{\theta})\right)^{-1} \\ &= \mathrm{Var}(\mathbf{BA}\,S(\mathbf{Y},\boldsymbol{\theta}))^{-1} = \mathbf{I},\end{aligned}$$

where the second equality holds because $\mathbf{B}$ is chosen so that $\mathrm{Var}(\mathbf{BA}S(\mathbf{Y},\boldsymbol{\theta})) = -E(\partial/\partial\tilde{\boldsymbol{\theta}}\mathbf{BA}\,S(\mathbf{Y},\boldsymbol{\theta}))$.

In practice, since the true parameter $\boldsymbol{\theta}$ is unknown, $\mathbf{A}$ and $\mathbf{B}$ are estimated using $\hat{\boldsymbol{\theta}}$, a consistent estimate of $\boldsymbol{\theta}$. Then, the $p \times p$ matrices $\mathbf{P}$ and $\mathbf{Q}$ are computed according to

$$\hat{\mathbf{P}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}_i(y_i,\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \text{ and } \hat{\mathbf{Q}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\psi}_i(y_i,\hat{\boldsymbol{\theta}})\,\boldsymbol{\psi}_i(y_i,\hat{\boldsymbol{\theta}})^{\mathrm{T}}\right\}. \tag{14}$$

The matrices $\mathbf{A}$ and $\mathbf{B}$ are then computed according to (13) and (12) with $\mathbf{P}$ and $\mathbf{Q}$ replaced by $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$, that is $\hat{\mathbf{A}} = (n \cdot \hat{\mathbf{P}}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{P}})^{-1/2}$ and $\hat{\mathbf{B}} = -\hat{\mathbf{A}}\hat{\mathbf{P}}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{A}}^{-1}$.

The following proposition shows the asymptotic validity of MCMB-**AB**. We follow the convention to say that $\{\psi_{ij}(.,\boldsymbol{\theta})\}$ is continuous in $\boldsymbol{\theta}$ uniformly in i,j if for any $\varepsilon>0$ there exists $\delta>0$ such that $\max_{i,j}|\psi_{ij}(.,\boldsymbol{\theta}_1) - \psi_{ij}(.,\boldsymbol{\theta}_2)| < \varepsilon$ whenever $||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|| < \delta$.

Proposition 1. *If $\psi_{ij}(Y_i,\boldsymbol{\theta})$ and $\partial\psi_{ij}(Y_i,\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ are continuous in $\boldsymbol{\theta}$ uniformly in i,j, and if the matrices $\mathbf{P}$ and $\mathbf{Q}$ of* (9) *exist and have finite limits $\mathbf{P}_\infty$ and $\mathbf{Q}_\infty$ as $n \to \infty$, then condition* (C3) *is satisfied when $\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta})$ is replaced by $\hat{\mathbf{B}}\hat{\mathbf{A}}\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta})$.*

Proof. Let $\boldsymbol{\theta}_0$ denote the true parameter. For convenience, let $\mathbf{A}^* = (\mathbf{P}_\infty\mathbf{Q}_\infty^{-1}\mathbf{P}_\infty)^{-1/2}$ and $\mathbf{B}^* = -\mathbf{A}^*\mathbf{P}_\infty\mathbf{Q}_\infty^{-1}\mathbf{A}^{*-1}$. Since $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$, and by continuity of $\boldsymbol{\psi}(\mathbf{Y},\boldsymbol{\theta})$ and its derivative with respect to $\boldsymbol{\theta}$, we have

$$\begin{aligned}&\hat{\mathbf{P}} - \tfrac{1}{n}\sum_{i=1}^{n}\tfrac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \to 0, \\ &\hat{\mathbf{Q}} - \tfrac{1}{n}\sum_{i=1}^{n}\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta}_0)\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta}_0)^{\mathrm{T}} \to 0.\end{aligned} \tag{15}$$

By the law of large numbers,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} - \mathbf{P} \to 0, \\ &\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta}_0)\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta}_0)^{\mathrm{T}} - \mathbf{Q} \to 0,\end{aligned}$$

and thus $\hat{\mathbf{P}} - \mathbf{P}_\infty \to 0$ and $\hat{\mathbf{Q}} - \mathbf{Q}_\infty \to 0$. Therefore, $\sqrt{n}\,\hat{\mathbf{A}} \to \mathbf{A}^*$ and $\hat{\mathbf{B}} \to \mathbf{B}^*$ as $n \to \infty$. By the construction of $\mathbf{A}^*$ and $\mathbf{B}^*$,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}(\mathbf{B}^*\mathbf{A}^*\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta}_0))(\mathbf{B}^*\mathbf{A}^*\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta}_0))^{\mathrm{T}} \quad \to \mathbf{B}^*\mathbf{A}^*\mathbf{Q}_\infty\mathbf{A}^{*\mathrm{T}}\mathbf{B}^{*\mathrm{T}}, \\ &-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\tilde{\boldsymbol{\theta}}}\mathbf{B}^*\mathbf{A}^*\boldsymbol{\psi}_i(Y_i,\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \to -\mathbf{B}^*\mathbf{A}^*\mathbf{P}_\infty.\end{aligned}$$

These limits also hold with $(\mathbf{A}^*\mathbf{B}^*)$ replaced by $\sqrt{(n)}\,\hat{\mathbf{A}}\hat{\mathbf{B}}$, and $\mathbf{D}(\boldsymbol{\theta}_0) = \mathbf{B}^*\mathbf{A}^*\mathbf{Q}_\infty\mathbf{A}^{*\mathrm{T}}\mathbf{B}^{*\mathrm{T}}$. □

Note that uniform continuity assumption in the proposition is only used to simplify the proof that leads to (15). In many cases, the results in (15) hold without such strong conditions. Under mild technical conditions, the MCMB-**AB** method remains valid if $\boldsymbol{\psi}$ has finitely many jumps, as in the quantile regression estimators. In such cases, the derivative of $\boldsymbol{\psi}$ in (14) needs to be replaced by derivative of the expected value of $\boldsymbol{\psi}$.

The MCMB-**AB** method extends the MCMB approach to general estimating equations by exploring the flexibility of redefining margins in a class of equivalent p-dimensional equations. In some cases, this would result in more complicated marginal equations, but one-dimensional equations are generally manageable even if a grid search algorithm is necessary to find a reliable root.

6. An MCMC perspective

The MCMB method shares one important feature with the Gibbs sampler, a popular MCMC algorithm, that is, both methods break up a high-dimensional problem into one-dimensional problems. In the asymptotic sense, one may view the update on the jth component of $\boldsymbol{\theta}$ as an attempt to draw from the conditional distribution of θ_j given the current values of $\boldsymbol{\theta}$ without the jth component, but the exact asymptotic equivalence here is difficult to specify. The differences between the MCMB method and the MCMC method are clearer, because the latter is a finite-sample procedure and the former is based on asymptotic normality of the estimators. As a consequence, only a short chain from the MCMB algorithm is usually needed to achieve convergence while a long chain is preferred for any MCMC method. This is a double-edged sword. Convergence of MCMC algorithms is often a source of complaint, but accuracy from any MCMC method generally increases with chain length. On the other hand, the MCMB algorithm starts with a consistent estimator of $\boldsymbol{\theta}$ and converges rather quickly, but error accumulation can occur when a long chain is run. In fact, we advice against using a long MCMB chain, because the accuracy of the MCMB algorithm is usually dominated by the normal approximation to the distribution of $\hat{\boldsymbol{\theta}}$.

A recent paper of Tian et al. (2007) provides a new perspective on the relationship between the MCMB method, the estimating bootstrap method (Hu and Zidek, 1995), and the MCMC approach to estimating equations. Tian et al. (2007) uses an MCMC algorithm to the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ based on a different **B**-transformation (from what is used in this paper) to achieve asymptotic normality of the estimating equation at the true parameter.

7. Illustrative examples

7.1. A simulation study

To illustrate the use of **A**- and **B**-transformations, we consider a simple two-parameter nonlinear regression model with a heteroscedastic error term

$$y_i = \theta_1 e^{-\theta_2 x_i} + e^{x_i} z_i \quad i = 1, \ldots, 500, \tag{16}$$

where $-1 < x_i < 2$, and $z_i \sim N(0, 1)$. The scatter plot of a typical data set generated from this model with $\theta_1 = 4$ and $\theta_2 = 1.5$ can be seen in Fig. 1. For this model **P** and **Q** are

$$\mathbf{P} = E\begin{pmatrix} -e^{-2x\theta_2} & xe^{-\theta_2 x}(\theta_1 e^{-\theta_2 x} - z) \\ xe^{-\theta_2 x}(\theta_1 e^{-\theta_2 x} - z) & -\theta_1 x^2 e^{-\theta_2 x}(\theta_1 e^{-\theta_2 x} - z) \end{pmatrix},$$

$$\mathbf{Q} = E\begin{pmatrix} z^2 e^{-2\theta_2 x} & -\theta_1 x z^2 e^{-2\theta_2 x} \\ -\theta_1 x z^2 e^{-2\theta_2 x} & \theta_1^2 x^2 z^2 e^{-2\theta_2 x} \end{pmatrix}.$$

We can empirically check whether condition (C3) holds by checking whether $\hat{\mathbf{P}}\hat{\mathbf{Q}}^{-1} \approx \mathbf{I}$. For the data set in Fig. 1 we find that

$$\hat{\mathbf{P}}\hat{\mathbf{Q}}^{-1} = \begin{pmatrix} -1.823 & -0.542 \\ -4.140 & -1.904 \end{pmatrix}$$

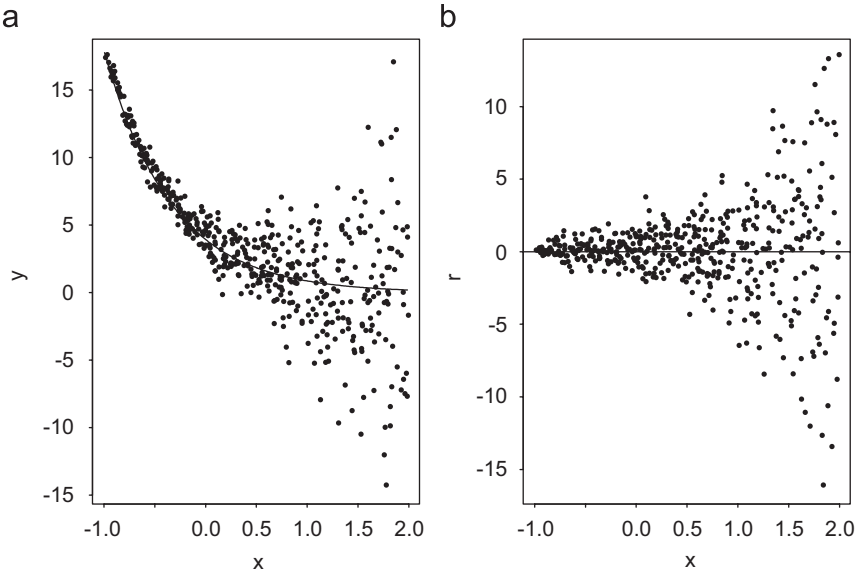

Fig. 1. (a) Scatter plot of the data in (16), (b) residuals from nls() vs. x.

with eigenvalues $\lambda_1 = -3.36, \lambda_2 = -0.37$. Recall that if the matrix is approximately equal to the identity matrix, the ratio of the largest and the smallest eigenvalues (also called the condition number of a matrix) should be close to 1. The condition number of in $\hat{\mathbf{P}}\hat{\mathbf{Q}}^{-1}$ is 9.22, suggesting that (C3) does not hold.

We compared the performance of the following methods: nonlinear least-squares regression (Bates and Watts, 1988) using `nls()` available in **R** (*NLS*), direct MCMB (*MCMB*), MCMB with the **A**-transformation only (*MCMB*-**A**), MCMB with the **B**-transformation only (*MCMB*-**B**), MCMB with the **A**- and **B**-transformations (*MCMB*-**AB**), the paired bootstrap (*PB*), and the Monte Carlo simulation (*MC*). All bootstrap estimates were based on $K = 500$ replications. In the Monte Carlo simulation $M = 500$ independent **y**'s were generated, and `nls()` was used to estimate the parameters.

The Markov chains $\hat{\boldsymbol{\theta}}^{(k)}$ produced by each of the four MCMB methods are plotted in Fig. 2. Applying MCMB directly generates Markov chains with some visible autocorrelation (a), suggesting that the **A**-transformation is needed. Panel (b) shows Markov chains produced MCMB with only the **A**-transformation. Although the resulting sequences look somewhat improved, clearly autocorrelation persists. Panel (c) shows MCMB chains generated by MCMB-**B**, which clearly performs substantially worse than direct MCMB or MCMB-**A**. Based on this and other models (not shown), we do not recommend using the **B**-transformation alone. Finally, panel (d) shows Markov chains resulting from MCMB-**AB**, where the sequences look completely random.

Mean and standard error estimates are reported in Table 1. Table 2 shows the corresponding coverage probabilities based on the asymptotic $(1-\alpha)100\%$ confidence intervals. Note that the estimated MCMB and MCMB-**A** standard errors are substantially smaller than in MC or PB methods, resulting in much lower than nominal coverage probabilities. However, using both transformations jointly dramatically improves the MCMB performance. MCMB-**AB** performs as well as the paired bootstrap, providing accurate coverage probabilities and parameter estimates close to the true $\boldsymbol{\theta}$. We also note the poor performance of the usual NLS estimators, which can be improved upon with either an appropriate transformation of the response or by using the robust sandwich variance–covariance estimator.

Fig. 2 and Table 1 both demonstrate that MCMB-**AB** improves the MCMB performance when both autocorrelation and heteroscedasticity are present. Clearly, MCMB-**B** had the worst performance of the four MCMB methods, and can introduce even worse autocorrelation in the MCMB chains than direct MCMB.

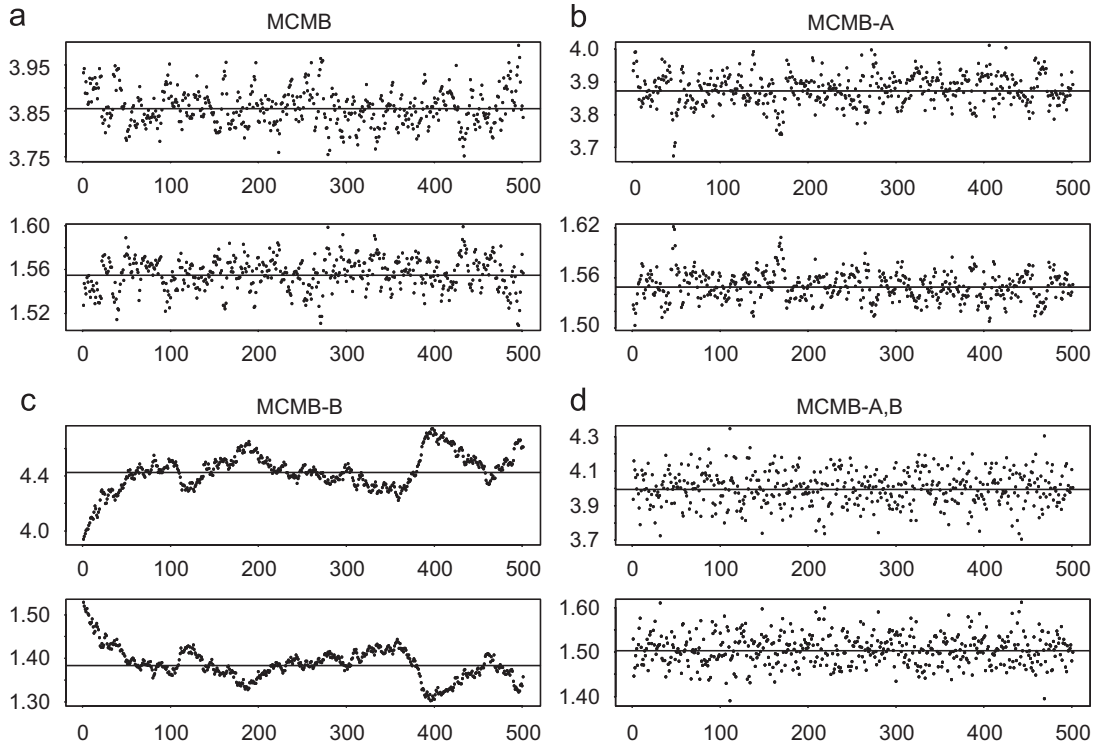

Fig. 2. (a) MCMB chains, (b) MCMB-**A**, (c) MCMB-**B**, (d) MCMB-**AB**. In each subfigure, the upper panel is for θ_1 and the lower panel is for θ_2.

Table 1
Parameter estimates and their standard errors

	$\hat{\theta}_1$	SE($\hat{\theta}_1$)	$\hat{\theta}_2$	SE($\hat{\theta}_2$)
NLS	3.8698	0.2017	1.5497	0.0726
MCMB	3.8551	0.0404	1.5549	0.0148
MCMB-**A**	3.8724	0.0475	1.5488	0.0171
MCMB-**B**	4.4246	0.1329	1.3838	0.0366
MCMB-**AB**	3.9943	0.0994	1.5029	0.0345
PB	3.8790	0.0984	1.5470	0.0346
MC	3.9978	0.1058	1.5001	0.0370

We also note that the confidence intervals can often be improved upon by using the bias-corrected and accelerated (BC_a) version of the percentile method for confidence interval estimation (Efron and Tibshirani, 1998), but it is yet to be determined whether BC_a will benefit the MCMB methods.

7.2. An example

We now demonstrate another property of the MCMB method in a relatively simple case: the variance of median regression estimates with leverage points in the data.

Suppose that the median of y given x is $\alpha + \beta x$, which is estimated by minimizing $\sum_{i=1}^{n} |y_i - \alpha - \beta x_i|$ for a given sample $\{(x_i, y_i)\} = \{(0,1),(1,1),(1,2),(2,3),(3,2),(10,10)\}$. This artificial data set includes a leverage point at (10,10). The usual paired bootstrap based on 100 bootstrap samples gave the standard error estimate of 0.36 for the estimate of β, while the MCMB method based on a chain of length 100 estimated the standard error to be 0.08. Both methods avoid the specification of likelihood.

Table 2
Coverage probability and average confidence interval length

	$\alpha = 0.1$				$\alpha = 0.05$			
	Coverage		Length		Coverage		Length	
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
MCMB	0.38	0.41	0.3121	0.1094	0.45	0.50	0.3719	0.1303
MCMB-A	0.47	0.48	0.3691	0.1272	0.57	0.54	0.4398	0.1515
MCMB-AB	0.89	0.89	0.8734	0.3134	0.93	0.95	1.0407	0.3734
PB	0.90	0.92	0.9075	0.3177	0.94	0.94	1.0814	0.3786
NLS	0.97	0.96	1.4862	0.5214	1.00	1.00	1.7710	0.6212

The significant difference in the results here can be easily attributed to the leverage point in the sample, but if the linear quantile model is true and the conditional distributions of y given x do not change much with x, the paired bootstrap inflates the variance. This problem with the usual bootstrap has been observed even in the univariate sample median studies, see for example, Ghosh et al. (1984) and Jimenez-Gamero et al. (1998). The MCMB method uses all the design points, so it captures the conditional regression analysis in a natural way. For algorithmic development of the MCMB method in quantile regression, we refer to Kocherginsky et al. (2005).

8. Conclusions

In this article we introduced a transformation of the estimating equations, called the **B**-transformation, which validates the MCMB for estimating equations that are not necessarily likelihood-based. We recommend that this transformation be used jointly with the **A**-transformation of the parameter space, which ''decorrelates'' the MCMB sequences.

The MCMB-**AB** approach for inference may be used as an alternative to the large-sample inference based on the robust variance–covariance matrix for any estimator given by a generalized estimating equation. The purpose of the MCMB approach (including the extensions discussed here) is not to improve accuracy over the usual bootstrap method. Instead, we focus on two distinctive features of the MCMB approach.

First, when the estimating equation is difficult to solve (especially when p is modestly large), the usual bootstrap is practically prohibitive. For example, some estimating equations are likely to have multiple roots, so finding the right root as a consistent estimator is challenging even in the original sample. In such cases, the MCMB method can be used by relying on a consistent estimator of $\boldsymbol{\theta}$ which is easier to compute but is not the same as the estimator defined by the estimating equation.

Second, the MCMB algorithm requires solving one-dimensional equations only, which is feasible even if the equation is highly nonlinear. The grid search might be the last resort. When multiple roots are present, we can always use the one that is closest to a consistent estimator.

However, consistent estimation of the **A**- and **B**-transformations may, in some cases, require as much as what is needed to estimate the robust variance–covariance matrix itself. The relative merits on the MCMB-**AB** method need to be established case by case. In the cases where **P** involves a nonparametric density function, as in the quantile regression estimation, the MCMB method has been shown to be robust even when the estimate of **P** itself and the estimate of the robust variance–covariance matrix are sensitive to the smoothing parameter needed for the density estimation. More details on the **A**-, **B**-transformations may be found at Kocherginsky (2003), the first author's dissertation at the University of Illinois.

References

Bates, D.M., Watts, D.G., 1988. Nonlinear Regression Analysis and its Applications. Wiley, New York, NY.
Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and their Application. Cambridge University Press, Cambridge.
Efron, B., Tibshirani, R.J., 1998. An Introduction to the Bootstrap. CRC Press LLC, Boca Raton, FL.

Ghosh, M., Parr, M.C., Singh, K., Babu, G.J., 1984. A note on bootstrapping the sample median. Ann. Statist. 12, 1130–1135.
He, X., Hu, F., 2002. Markov chain marginal bootstrap. J. Amer. Statist. Assoc. 97 (459), 783–795.
Hu, F., Zidek, J.V., 1995. A Bootstrap based on the estimating equations of the linear model. Biometrika 82, 263–275.
Huber, P.J., 1964. Robust estimation of a location parameter. Ann. Math. Statist. 35, 73–101.
Jimenez-Gamero, M.D., Munoz-Garcia, J., Pino-Mejias, R., 1998. Reduced bootstrap for the median. Statist. Sinica 14, 1179–1198.
Kocherginsky, M., 2003. Extensions of Markov chain marginal bootstrap. Ph.D. Thesis, University of Illinois at Urbana-Champaign.
Kocherginsky, M., He, X., Mu, Y., 2005. Practical confidence intervals for regression quantiles. J. Comput. Graphical Statist. 14, 41–55.
Koenker, R.W., Bassett, G.J., 1978. Regression quantiles. Econometrica 46, 33–50.
Tian, L., Liu, J.S., Wei, L.J., 2007. Implementation of estimating-equation based inference procedures with MCMC samplers. J. Amer. Statist. Assoc., in press.